

1/18

1 accacaaaca cuccaguuug uuacacuccg cuaggaaugc uccuggagca
51 cccccccuag cagggcgugg gggauuuccc cugcccgucu gcagaagggg
101 ggagccaacc accuuaguau guaggcggcg ggacucauga cgcucgcgug
151 augacaagcg ccaagcuuga cuuggauggc ccugaugggc guucaugggu
201 ucgguggugg uggcgcuuaa ggcagccucc acgcccacca ccucccagau
251 agagcgggcg cacuguaggg aagaccgggg accggucacu accaaggacg
301 cagaccucuu uuugaguauc acgccuccgg aaguaguugg gcaagcccac
351 cuauaugugu ugggaugguu gggguuagcc auccauaccg uacugccuga
401 uaggguccuu gcgagggggu cugggagucu cguagaccgu agcacAUGCC
451 UGUUAUUUCU ACUCAACAA GUCCUGUACC UGCGCCCAGA ACGCGCAAGA
501 ACAAGCAGAC Ggggcgcgcc augauugaac aagauggauu gcacgcaggu
551 ucuccggccg cuugggugga gaggcuauc ggcuangacu gggcacaaca
601 gacaauccgc ugcucugaug ccgccguguu ccggcuguca gcgcaggggc
651 gcccgguuu uuuugucaag accgaccugu ccggugcccu gaaugaacug
701 caggacgagg cagcgcggu aucguggcug gccacgacgg gcguuccuug
751 cgcagcugug cucgacguug ucacugaagc gggaaggac uggcugcuau
801 ugggcgaagu gccggggcag gaucuccugu caucucaccu ugcuccugcc
851 gagaagguau ccaucauggc ugaugcaug cggcggcugc auacgcuuga
901 uccggcuacc ugcccauucg accaccaagc gaaacaucgc aucgagcgag
951 cacguacucg gauggaagcc ggucuugucg aucaggauga ucuggacgag
1001 gagcaucagg ggcucgcgcc agccgaacug uucgccaggc ucaaggcgcg
1051 caugcccgcac ggcgaggau ucgucgugac ccauggcgau gccugcuugc
1101 cgaauaucau gguggaaaau ggccgcuuuu cuggauucau cgacuguggc
1151 cggcugggug uggcggaccg cuaucaggac auagcguugg cuacccguga
1201 uauugcugag gagcuuggcg gcgaaugggc ugaccgcuuc cucgugcuuu
1251 acgguaucgc cgcucccgau ucgcagcgca ucgccuucua ucgccuucuu
1301 gacgaguucu ucugaguuaa aacagaccac aacgguuucc cucuagcggg
1351 aucaauuccg cccucuccc ucccccccc cuaacguuac uggccgaagc
1401 cgcuuggaau aaggccggug ugcguuuguc uauauguuau uuuccaccu

FIG. 1A

2/18

1451 auugccgucu uuuggcaaug ugagggcccg gaaaccuggc ccugucuucu
1501 ugacgagcau uccuaggggu cuuuccccuc ucgccaaagg aaugcaaggu
1551 cuguugaaug ucgugaagga agcaguuccu cuggaagcuu cuugaagaca
1601 aacaacgucu guagcgaccc uuugcaggca gcggaacccc ccaccuggcg
1651 acaggugccu cugcggccaa aagccacgug uauaagauac accugcaaag
1701 gcggcacaac cccagugcca cguugugagu uggauaguug uggaaagagu
1751 caaauggcuc uccucaagcg uauucaaaa ggggcugaag gaugcccaga
1801 agguacccca uuguauggga ucugaucugg ggccucggug cacaugcuuu
1851 acauguguuu agucgagguu aaaaaacguc uaggccccc gaaccacggg
1901 gacgugguuu uccuuugaaa aacacgauaa uaccauggca ccuuuuacgc
1951 ugcagugucu cucugaacgu ggcacgcugu cagcgauggc aguggucaug
2001 acugguauag acccccgaac uuggacugga acuaucuuca gauuaggauc
2051 ucuggccacu agcuacaugg gauuuguuug ugacaacgug uuguauacug
2101 cucaccaugg cagcaagggg cgccgguugg cucauccac aggcuccaua
2151 cacccaauaa ccguugacgc ggcuaaugac caggacauca aucaaccacc
2201 auguggagcu gggucccuua cucggugcuc uugcggggag accaaggggu
2251 aucugguaac acgacugggg ucauugguug aggucaauaa auccgaugac
2301 ccuuauuggu gugugugcgg ggcccuuccc auggcuguug ccaagggguu
2351 uucaggugcc ccgauucugu gcuccuccgg gcauguuauu ggganguuca
2401 ccgcugcuag aaauucuggc gguucaguca gccagauuag gguuagggcg
2451 uuggugugug cuggauacca ucccaguac acagcacaug ccacucuuga
2501 uacaaaaccu acugugccua acgaguauuc agugcaaaau uuaauugccc
2551 ccacuggcag cggcaaguca accaaaauac cacuuucuaa caugcaggag
2601 aaguaugagg ucuugguccu aaaucccagu guggcuacaa cagcaucaau
2651 gccaaaguac augcacgcga cguacggcgu gaauccaaau ugcuauuuua
2701 auggcaaau uaccaacaca ggggcuucac uuacguacag cacauauggc
2751 auguaccuga ccggagcaug uucccggaac uaugauguaa ucauuuguga
2801 cgaaugccau gcuaccgaug caaccaccgu guugggcauu ggaaaggucc
2851 uaaccgaagc uccauccaaa aauguuaggg uagugguucu ugccacggcu

FIG. 1B

3/18

2901 accccccccug gaguaauccc uacaccacau gccaacauaa cugagauuca
2951 auuaaccgau gaaggcacua ucccuuuca uggaaaaaag auuaaggagg
3001 aaaaucugaa gaaagggaga caccuuaucu uugaggcuac caaaaaacac
3051 ugugaugagc uugcuaacga guuagcucga aagggaaaua cagcugucuc
3101 uuacuauagg ggaugugaca ucucaaaaau cccugagggc gacuguguag
3151 uaguugccac ugaugccuug uguacagggu acacugguga cuuugauucc
3201 guguaugacu gcagccucau gguagaaggc acaugccaug uugaccuuga
3251 cccuacuuc accaugggug uucgugugug cgggguuuca gcaauaguua
3301 aaggccagcg uagggggccgc acaggccgug ggagagcugg cauauacuac
3351 uauguagacg ggaguuguac cccuucgggu augguuccug aaugcaacau
3401 uguugaagcc uucgacgcag ccaaggcaug guaugguuug ucaucaacag
3451 aagcucaaac uauucuggac accuaucgca cccaaccugg guuaccugcg
3501 auaggagcaa auuuggacga gugggcugau cucuuucaa uggucaaccc
3551 cgaaccuua uuugucaaua cugcaaaaag aacugcugac aaauauguuu
3601 uguugacugc agcccaacua caacuguguc aucaguaugg cuaugcugcu
3651 cccaugacg caccacggug gcagggagcc cggcuuggga aaaaaccuug
3701 uggggguucug uggcgcuugg acggcgugcga cgccuguccu ggcccagagc
3751 ccagcgaggu gaccagauac caaauugucu ucacugaagu caauacuucu
3801 gggacagccg cacucgcugu uggcguugga guggcuauagg cuuaucuagc
3851 cauugacacu uuuggcgcca cuugugugcg gcguugcugg ucuaauacau
3901 cagucccuac cggugcuacu gucgccccag ugguugacga agaagaauc
3951 guggaggagu gugcaucau cauucccuug gaggccaugg uugcugcaau
4001 ugacaagcug aagaguacaa ucaccacaac uaguccuuc acauuggaaa
4051 ccgcccuga aaaacuuaac accuuucuuug ggccucaugc agcuacauc
4101 cuugcuauc uagaguauug cuguggcuua gucacuuuac cugacaauc
4151 cuuugcauc ugcguguuug cuuucuuugc ggguaauacu accccacuac
4201 cucacaagau caaauguuc cugucauuau uuggaggcg aaugcgucc
4251 aagcuuacag acgcuagagg cgcacuggcg uucaugaugg cggggcgucg
4301 gggaacagcu cuugguacau ggacaucggu ggguuuuguc uuugacaugc

FIG. 1C

4/18

4351 uaggcggcua ugcugccgcc ucauccacug cuugcuugac auuuaaaugc
4401 uugaugggug aguggccac uauggaucag cuugcugguu uagucuacuc
4451 cgcguucaac ccggccgcag gaguuguggg cguuuuguca gcuugugcaa
4501 uguuugcuuu gacaacagca gggccagauc acuggcccaa cagacuucuu
4551 acuaugcuug cuaggagcaa cacuguaugu aaugaguacu uuauugccac
4601 ucgugacauc cgcaggaaga uacugggcau ucuggaggca ucuaccccu
4651 ggagugucau aucagcuugc auccguuggc uccacacccc gacggaggau
4701 gauugcggcc ucauugcuug gggucuaagag auuuggcagu acgugugcaa
4751 uuucuuugug auuugcuuaa auguccuaa agcuggaguu cagagcaugg
4801 uuaacauucc ugguuguccu uucuacagcu gccagaaggg guacaagggc
4851 ccugggauug gaucaggauu gcuccaagca cgcuguccau gcgguugcuga
4901 acucaucuuu ucuguugaga augguuuugc aaaacuuuac aaaggaccca
4951 gaacuuguuc aaauuacugg agaggggcug uuccagucaa cgcuaaggcug
5001 ugugggucgg cuagaccgga cccaacugau uggacuaguc uugucgucaa
5051 uuauggcguu agggacuacu guaaaauaga gaaaauuggga gaucacauuc
5101 uuguuacagc aguauccucu ccaaauugcu guuucaccca ggugcccca
5151 accuugagag cugcaguggc cguggacggc guacagguuc aguguuauuc
5201 aggugagccc aaaacuccuu ggacgacauc ugcuugcugu uacgguccug
5251 acgguaaggg uaaaacuguu aagcuucccu uccgcguuga cggucacaca
5301 ccuggugugc gcaugcaacu uaauuugcgu gaugcacuug agacaaauga
5351 cuguauuucc acaacaaca cuccuaguga ugaagccgca guguccgcuc
5401 uuguuuucaa acaggaguug cggcguacaa accaaugcu ugaggcaauu
5451 ucagcuggcg uugacaccac caaacugcca gccccucca ucgaagaggu
5501 agugguaaga aagcgccagu uccgggcaag aacugguucg cuuaccuugc
5551 cccccuccc gagauccguc ccaggagugu cauguccuga aagccugcaa
5601 cgaagugacc cguuagaagg uccuucacaa cuccuccuu caccaccugu
5651 ucuacaguug gccaugccga ugccccuguu gggagcgggu gaguguaacc
5701 cuuucacugc aaugggaugu gcaugaccg aaacaggcgg aggccugau
5751 gauuuaccca guuaccucc caaaaaggag gucucugaau ggucagacga

FIG. 1D

5/18

5801 aaguugguca acggcuacaa ccgcuuccag cuacguuacu ggccccccgu
5851 acccuaagau acggggaaag gauuccacuc agucagcccc cgccaaacgg
5901 ccuacaaaaa agaaguuggg aaagagugag uuuucgugca gcaugagcua
5951 cacuuggacc gacgugauua gcuuacaaac ugcuuuuaaa guucugucug
6001 caacucgggc caucacuagu gguuuccuca aacaaagauc auuggugua
6051 gugacugagc cgcgggaugc ggagcuuaga aaacaaaaag ucacuaauaa
6101 uagacaaccu cuguuccccc caucauacca caagcaagug agauuggcua
6151 aggagaaagc uucaaaaguu gucgguguca ugugggacua ugaugaagua
6201 gcagcucaca cgcccucuaa gucugcuaag ucccacauca cuggccuucg
6251 gggcacugau guucguucug gagcagcccg caaggcuguu cuggacuugc
6301 agaagugugu cgaggcaggu gagauaccga gucauuauucg gcaaacagug
6351 auaguuccaa aggaggaggu cuucgugaag acccccaga aaccaacaaa
6401 gaaacccccca aggcucaucu cguacccccca ccuugaaaug agauguguug
6451 agaagaugua cuacggucag guugcuccug acguaguuaa agcugucaug
6501 ggagaugcgu acggguuugu agauccacgu acccguguca agcgucuguu
6551 gucgaugugg ucacccgaug cagucggagc cacaugcgau acaguguguu
6601 uugacaguac caucacaccc gaggauauca ugguggagac agacaucuac
6651 ucagcagcua aacucaguga ccaacaccga gcuggcauuc acaccauugc
6701 gaggcaguua uacgcuggag gaccgaugau cgcuuaugau ggccgagaga
6751 ucggauaucg uagguguagg ucuuccggcg ucuauacuac cucaaguucc
6801 aacaguuga ccugcuggcu gaagguaau gcugcagccg aacaggcugg
6851 caugaagaac ccucgcuucc uuauuugcgg cgaugauugc accguaauu
6901 ggaaaagcgc cggagcagau gcagacaaac aagcaaugcg ugucuuugcu
6951 agcuggauga aggugauggg ugcaccacaa gauugugugc cucaacccaa
7001 auacaguug gaagaauuaa caucaugcuc aucaaauguu accucuggaa
7051 uuacaaaaag uggcaagccu uacuacuuc uuacaagaga uccucguauc
7101 ccccuuggca ggugcucugc cgagggucug ggauacaacc ccagugcugc
7151 guggauuggg uaucuaauac aucacuacc auguuugugg guuagccgug
7201 uguuggcugu ccuuuucaug gagcagaugc ucuuugagga caaacuuccc

FIG. 1E

6/18

7251 gagacuguga ccuuugacug guaugggaaa aauuauacgg ugccuguaga
7301 agaucugccc agcaucauug cuggugugca cgguaauugag gcuuucucgg
7351 uggugcgcua caccaacgcu gagauccuca gaguuuccca aucacuaaca
7401 gacaugacca ugcccccccu gcgagccugg cgaaagaaag ccagggcggu
7451 ccucgccagc gccaaagaggc guggcgaggc acacgcaaaa ugggcucgcu
7501 uccuucucug gcaugcuaca ucuagaccuc uaccagauuu ggauaagacg
7551 agcguggcuc gguacaccac uuucaauuau ugugauguuu acuccccgga
7601 ggggggaugug uuuguuacac cacagagaag auugcagaag uuucuuguga
7651 aguauuuggc ugucauuguu uuugcccuag ggcucauugc uguuggauua
7701 gccaucagcu gaacccccaa auucaaaaau aacuaacagu uuuuuuuuuu
7751 uuuuuuuuuu agggcagcgg caacagggga gaccccgggc uuaacgaccc
7801 cgccgaugug aguugggcga ccauggugga ucagaaccgu uucgggugaa
7851 gccauggucu gaaggggaug acgucccuuc uggcucaucc aaaaaaccg
7901 ucucgggugg gugaggaguc cuggcugugu gggaagcagu caguauaaau
7951 cccgucgugu guggugacgc cucacgacgu acuuguccgc ugugcagagc
8001 guaguaccaa gggcugcacc ccgguuuuug uuccaagcgg agggcaaccc
8051 ccgcuuggaa uuaaaaacu

FIG. 1F

7/18

1 ACCACAAACA CTCCAGTTTG TTACACTCCG CTAGGAATGC TCCTGGAGCA
51 CCCCCCTAG CAGGGCGTGG GGGATTTCCC CTGCCCCGTCT GCAGAAGGGT
101 GGAGCCAACC ACCTTAGTAT GTAGGCGGCG GGACTCATGA CGCTCGCGTG
151 ATGACAAGCG CCAAGCTTGA CTTGGATGGC CCTGATGGGC GTTCATGGGT
201 TCGGTGGTGG TGGCGCTTTA GGCAGCCTCC ACGCCCACCA CCTCCCAGAT
251 AGAGCGGCGG CACTGTAGGG AAGACCGGGG ACCGGTCACT ACCAAGGACG
301 CAGACCTCTT TTTGAGTATC ACGCCTCCGG AAGTAGTTGG GCAAGCCCAC
351 CTATATGTGT TGGGATGGTT GGGGTTAGCC ATCCATACCG TACTGCCTGA
401 TAGGGTCCTT GCGAGGGGAT CTGGGAGTCT CGTAGACCGT AGCACATGCC
451 TGTTATTTCT ACTCAAACAA GTCCTGTACC TGCGCCCAGA ACGCGCAAGA
501 ACAAGCAGAC GCAGGCTTCA TATCCTGTGT CCATTAAAC ATCTGTTGAA
551 AGGGGACAAC GAGCAAAGCG CAAAGTCCAG CGCGATGCTC GGCCTCGTAA
601 TTACAAAATT GCTGGTATCC ATGATGGCTT GCAGACATTG GCTCAGGCTG
651 CTTTGCCAGC TCATGGTTGG GGACGCCAAG ACCCTCGCCA TAAGTCTCGC
701 AATCTTGGA TCCTTCTGGA TTACCCTTTG GGGTGGATTG GTGATGTTAC
751 AACTCACACA CCTCTAGTAG GCCCGCTGGT GGCAGGAGCG GTCGTTCGAC
801 CAGTCTGCCA GATAGTACGC TTGCTGGAGG ATGGAGTCAA CTGGGCTACT
851 GGTGTTGTTG GTGTCCACCT TTTTGTGGTA TGTCTGCTAT CTTTGGCCTG
901 TCCCTGTAGT GGGGCGCGGG TCACTGACCC AGACACAAAT ACCACAATCC
951 TGACCAATTG CTGCCAGCGT AATCAGGTTA TCTATTGTTT TCCTTCCACT
1001 TGTCTACACG AGCCTGGTTG TGTGATCTGT GCGGACGAGT GCTGGGTTCC
1051 CGCCAATCCG TACATCTCAC ACCCTTCCAA TTGGACTGGC ACGGACTCCT
1101 TCTTGGCTGA CCACATTGAT TTTGTTATGG GCGCTCTTGT GACCTGTGAC
1151 GCCCTTGACA TTGGTGAGTT GTGTGGTGCG TGTGTATTAG TCGGTGACTG
1201 GCTTGTCAGG CACTGGCTTA TTCACATAGA CCTCAATGAA ACTGGTACTT
1251 GTTACCTGGA AGTGCCCACT GGAATAGATC CTGGGTTTCCT AGGGTTTATC
1301 GGGTGGATGG CCGGCAAGGT CGAGGCTGTC ATCTTCTTGA CCAAACCTGGC
1351 TTCACAAGTA CCATACGCTA TTGCGACTAT GTTTAGCAGT GTACACTACC
1401 TGGCGGTTGG CGCTCTGATC TACTATGCCT CTCGGGGCAA GTGGTATCAG

FIG. 2A

8/18

1451 TTGCTCCTAG CGCTTATGCT TTACATAGAA GCGACCTCTG GAAACCCCAT
1501 CAGGGTGCCC ACTGGATGCT CAATAGCTGA GTTTTGCTCG CCTTTGATGA
1551 TACCATGTCC TTGCCACTCT TATTTGAGTG AGAATGTGTC AGAAGTCATT
1601 TGTTACAGTC CAAAGTGGAC CAGGCCTGTC ACTCTAGAGT ATAACAACCT
1651 CATATCTTGG TACCCCTATA CAATCCCTGG TCGGAGGGGA TGTATGGTTA
1701 AATTCAAAAA TAACACATGG GGTGTGTGCC GTATTCGCAA TGTGCCATCG
1751 TACTGCACTA TGGGCACTGA TGCAGTGTGG AACGACACTC GCAACACTTA
1801 CGAAGCATGC GGTGTAACAC CATGGCTAAC AACC GCATGG CACAACGGCT
1851 CAGCCCTGAA ATTGGCTATA TTACAATACC CTGGGTCTAA AGAAATGTTT
1901 AAACCTCATA ATTGGATGTC AGGCCATTTG TATTTTGAGG GATCAGATAC
1951 CCCTATAGTT TACTTTTATG ACCCTGTGAA TTCCACTCTC CTACCACCGG
2001 AGAGGTGGGC TAGGTTGCCC GGTACCCAC CTGTGGTACG TGGTTCTTGG
2051 TTACAGGTTT CGCAAGGGTT TTACAGTGAT GTGAAAGACC TAGCCACAGG
2101 ATTGATCACC AAAGACAAAG CCTGGAAAAA TTATCAGGTC TTATATTCCG
2151 CCACGGGTGC TTTGTCTCTT ACGGGAGTTA CCACCAAGGC CGTGGTGCTA
2201 ATTCTGTTGG GGTGTGTGG CAGCAAGTAT CTTATTTTAG CCTACCTCTG
2251 TTA CTGTGCC CTTTGT TTTG GCGCGCTTC TGGTTACCCT TTGCGTCTG
2301 TGCTCCCATC CCAGTCGTAT CTCCAAGCTG GCTGGGATGT TTTGTCTAAA
2351 GCTCAAGTAG CTCCTTTTGC TTTGATTTTC TTCATCTGTT GCTATCTCCG
2401 CTGCAGGCTA CGTTATGCTG CCCTTTTAGG GTTTGTGCCC ATGGCTGCGG
2451 GCTTGCCCCCT AACTTTCCTT GTTGCAGCAG CTGCTGCCCA ACCAGATTAT
2501 GACTGGTGGG TGCGACTGCT AGTGGCAGGG TTAGTTTTGT GGGCCGGCCG
2551 TGACCGTGGT CACCGCATAG CTCTGCTTGT AGGTCCTTGG CCTCTGGTAG
2601 CGCTTTTAAC CCTCTTGCAT TTGGTTACGC CTGCTTCAGC TTTTGACACC
2651 GAGATAATTG GAGGGCTGAC AATACCACCT GTAGTAGCAT TAGTTGTCAT
2701 GTCTCGTTTT GGCTTCTTTG CTCACTTGTT ACCTCGCTGT GCTTTAGTTA
2751 ACTCCTATCT TTGGCAACGT TGGGAGAATT GGTTTTGGAA CGTTACACTA
2801 AGACCGGAGA GGTTTTTCT TGTGCTGGTT TGTTTCCCCG GTGCGACATA
2851 TGACGCGCTG GTGACTTTCT GTGTGTGTCA CGTAGCTCTC CTATGTTTAA

FIG. 2B

9/18

2901 CATCCAGTGC AGCATCGTTC TTTGGGACTG ACTCTAGGGT TAGGGCCCAT
2951 AGAATGTTGG TCGTCTCGG AAAGTGCCAT GCTTGGTATT CTCATTATGT
3001 TCTTAAGTTT TTCCTCTTAG TGTTTGGTGA GAATGGTGTG TTTTCTATA
3051 AGCACTTGCA TGGTGATGTC TTGCCTAATG ATTTTGCCTC GAAACTACCA
3101 TTGCAAGAGC CATTTTTCCC TTTTGAAGGC AAGGCAAGGG TCTATAGGAA
3151 TGAAGGAAGA CGCTTGGCGT GTGGGGACAC GGTTGATGGT TTGCCCGTTG
3201 TTGCGCGTCT CGGCGACCTT GTTTTCGCAG GGTTAGCTAT GCCGCCAGAT
3251 GGGTGGGCCA TTACCGCACC TTTTACGCTG CAGTGTCTCT CTGAACGTGG
3301 CACGCTGTCA GCGATGGCAG TGGTCATGAC TGGTATAGAC CCCCGAACCTT
3351 GGACTGGAAC TATCTTCAGA TTAGGATCTC TGGCCACTAG CTACATGGGA
3401 TTTGTTTGTG ACAACGTGTT GTATACTGCT CACCATGGCA GCAAGGGGCG
3451 CCGGTTGGCT CATCCCACAG GCTCCATACA CCCAATAACC GTTGACGCGG
3501 CTAATGACCA GGACATCTAT CAACCACCAT GTGGAGCTGG GTCCCTTACT
3551 CGGTGCTCTT GCGGGGAGAC CAAGGGGTAT CTGGTAACAC GACTGGGGTC
3601 ATTGGTTGAG GTCAATAAAT CCGATGACCC TTATTGGTGT GTGTGCGGGG
3651 CCCTTCCCAT GGCTGTTGCC AAGGGTTCTT CAGGTGCCCC GATTCTGTGC
3701 TCCTCCGGG ATGTTATTGG GATGTTACCC GCTGCTAGAA ATTCTGGCGG
3751 TTCAGTCAGC CAGATTAGGG TTAGGCCGTT GGTGTGTGCT GGATACCATC
3801 CCCAGTACAC AGCACATGCC ACTCTTGATA CAAAACCTAC TGTGCCTAAC
3851 GAGTATTAGC TGCAAAATTT AATTGCCCC ACTGGCAGCG GCAAGTCAAC
3901 CAAATTACCA CTTTCTTACA TGCAGGAGAA GTATGAGGTC TTGGTCCTAA
3951 ATCCCAGTGT GGCTACAACA GCATCAATGC CAAAGTACAT GCACGCGACG
4001 TACGGCGTGA ATCCAAATTG CTATTTTAAT GGCAAATGTA CCAACACAGG
4051 GGCTTCACTT ACGTACAGCA CATATGGCAT GTACCTGACC GGAGCATGTT
4101 CCCGGAAC TAATGTAATC ATTTGTGACG AATGCCATGC TACCGATGCA
4151 ACCACCGTGT TGGGCATTGG AAAGGTCCTA ACCGAAGCTC CATCCAAAAA
4201 TGTTAGGCTA GTGGTTCTTG CCACGGCTAC CCCCCCTGGA GTAATCCCTA
4251 CACCACATGC CAACATAACT GAGATTCAAT TAACCGATGA AGGCACTATC
4301 CCCTTTCATG GAAAAAAGAT TAAGGAGGAA AATCTGAAGA AAGGGAGACA

FIG. 2C

10/18

4351 CCTTATCTTT GAGGCTACCA AAAAACACTG TGATGAGCTT GCTAACGAGT
4401 TAGCTCGAAA GGGAATAACA GCTGTCTCTT ACTATAGGGG ATGTGACATC
4451 TCAAAAATCC CTGAGGGCGA CTGTGTAGTA GTTGCCACTG ATGCCTTGTG
4501 TACAGGGTAC ACTGGTGACT TTGATTCCGT GTATGACTGC AGCCTCATGG
4551 TAGAAGGCAC ATGCCATGTT GACCTTGACC CTACTTTCAC CATGGGTGTT
4601 CGTGTGTGCG GGGTTTCAGC AATAGTTAAA GGCCAGCGTA GGGGCCGCAC
4651 AGGCCGTGGG AGAGCTGGCA TATACTACTA TGTAGACGGG AGTTGTACCC
4701 CTTCGGGTAT GGTTCCCTGAA TGCAACATTG TTGAAGCCTT CGACGCAGCC
4751 AAGGCATGGT ATGGTTTGTC ATCAACAGAA GCTCAAAC TA TTCTGGACAC
4801 CTATCGCACC CAACCTGGGT TACCTGCGAT AGGAGCAAAT TTGGACGAGT
4851 GGGCTGATCT CTTTTCAATG GTCAACCCCG AACCTTCATT TGTCAATACT
4901 GCAAAAAGAA CTGCTGACAA TTATGTTTTG TTGACTGCAG CCCAACTACA
4951 ACTGTGTCAT CAGTATGGCT ATGCTGCTCC CAATGACGCA CCACGGTGGC
5001 AGGGAGCCCG GCTTGGGAAA AAACCTTG TG GGGTTCTGTG GCGCTTGGAC
5051 GGCCTGACG CCTGTCCTGG CCCAGAGCCC AGCGAGGTGA CCAGATACCA
5101 AATGTGCTTC ACTGAAGTCA ATACTTCTGG GACAGCCGCA CTCGCTGT TG
5151 GCGTTGGAGT GGCTATGGCT TATCTAGCCA TTGACACTTT TGGCGCCACT
5201 TGTGTGCGGC GTTGCTGGTC TATTACATCA GTCCCTACCG GTGCTACTGT
5251 CGCCCCAGTG GTTGACGAAG AAGAAATCGT GGAGGAGTGT GCATCATTCA
5301 TTCCCTTGGA GGCCATGGTT GCTGCAATTG ACAAGCTGAA GAGTACAATC
5351 ACCACAAC TA GTCCTTTCAC ATTGGAAACC GCCCTTGAAA AACTTAACAC
5401 CTTTCTTGGG CCTCATGCAG CTACAATCCT TGCTATCATA GAGTATTGCT
5451 GTGGCTTAGT CACTTTACCT GACAATCCCT TTGCATCATG CGTGTTTGCT
5501 TTCATTGCGG GTATTACTAC CCCACTACCT CACAAGATCA AAATGTT CCT
5551 GTCATTATTT GGAGGCGCAA TTGCGTCCAA GCTTACAGAC GCTAGAGGCG
5601 CACTGGCGTT CATGATGGCC GGGGCTGCGG GAACAGCTCT TGGTACATGG
5651 ACATCGGTGG GTTTTGTCTT TGACATGCTA GGCGGCTATG CTGCCGCCTC
5701 ATCCACTGCT TGCTTGACAT TTAAATGCTT GATGGGTGAG TGGCCCACTA
5751 TGGATCAGCT TGCTGGTTTA GTCTACTCCG CGTTCAACCC GGCCGCAGGA

FIG. 2D

11/18

5801 GTTGTGGGCG TTTTGTGAGC TTGTGCAATG TTTGCTTTGA CAACAGCAGG
5851 GCCAGATCAC TGGCCCAACA GACTTCTTAC TATGCTTGCT AGGAGCAACA
5901 CTGTATGTAA TGAGTACTTT ATTGCCACTC GTGACATCCG CAGGAAGATA
5951 CTGGGCATTC TGGAGGCATC TACCCCTGG AGTGTCATAT CAGCTTGCAT
6001 CCGTTGGCTC CACACCCCGA CGGAGGATGA TTGCGGCCTC ATTGCTTGGG
6051 GTCTAGAGAT TTGGCAGTAC GTGTGCAATT TCTTTGTGAT TTGCTTTAAT
6101 GTCCTTAAAG CTGGAGTTCA GAGCATGGTT AACATTCCTG GTTGTCTTTT
6151 CTACAGCTGC CAGAAGGGGT ACAAGGGCCC CTGGATTGGA TCAGGTATGC
6201 TCCAAGCACG CTGTCCATGC GGTGCTGAAC TCATCTTTTC TGTTGAGAAT
6251 GGTTTTGCAA AACTTTACAA AGGACCCAGA ACTTGTTCOA ATTACTGGAG
6301 AGGGGCTGTT CCAGTCAACG CTAGGCTGTG TGGGTCGGCT AGACCGGACC
6351 CAACTGATTG GACTAGTCTT GTCGTCAATT ATGGCGTTAG GGACTACTGT
6401 AAATATGAGA AATTGGGAGA TCACATCTTT GTTACAGCAG TATCCTCTCC
6451 AAATGTCTGT TTCACCCAGG TGCCCCAAC CTTGAGAGCT GCAGTGGCCG
6501 TGGACGGCGT ACAGGTTTCA GTTTATCTAG GTGAGCCCAA AACTCCTTGG
6551 ACGACATCTG CTTGCTGTTA CGGTCCTGAC GGTAAGGGTA AAAGTGTAA
6601 GCTTCCCTTC CGCGTTGACG GTCACACACC TGGTGTGCGC ATGCAACTTA
6651 ATTTGCGTGA TGCACCTGAG ACAAATGACT GTAATTCCAC AAACAACACT
6701 CCTAGTGATG AAGCCGCAGT GTCCGCTCTT GTTTTCAAAC AGGAGTTGCG
6751 GCGTACAAAC CAATTGCTTG AGGCAATTTT AGCTGGCGTT GACACCACCA
6801 AACTGCCAGC CCCCTCCATC GAAGAGGTAG TGGTAAGAAA GCGCCAGTTC
6851 CGGGCAAGAA CTGGTTCGCT TACCTTGCCC CCCCCTCCGA GATCCGTCCC
6901 AGGAGTGTC TGTCTGAAA GCCTGCAACG AAGTGACCCG TTAGAAGGTC
6951 CTTCAAACCT CCTCCTTCA CCACCTGTTC TACAGTTGGC CATGCCGATG
7001 CCCCTGTTGG GAGCGGGTGA GTGTAACCCT TTCACTGCAA TTGGATGTGC
7051 AATGACCGAA ACAGGCGGAG GCCCTGATGA TTTACCCAGT TACCCTCCCA
7101 AAAAGGAGGT CTCTGAATGG TCAGACGAAA GTTGGTCAAC GGCTACAACC
7151 GCTTCCAGCT ACGTTACTGG CCCCCGTAC CCTAAGATAC GGGGAAAGGA
7201 TTCCACTCAG TCAGCCCCCG CCAAACGGCC TACAAAAAAG AAGTTGGGAA

FIG. 2E

12/18

7251 AGAGTGAGTT TTCGTGCAGC ATGAGCTACA CTTGGACCGA CGTGATTAGC
7301 TTCAAAACTG CTTCTAAAGT TCTGTCTGCA ACTCGGGCCA TCACTAGTGG
7351 TTTCCCTCAA CAAAGATCAT TGGTGTATGT GACTGAGCCG CGGGATGCGG
7401 AGCTTAGAAA ACAAAAAGTC ACTATTAATA GACAACCTCT GTTCCCCCA
7451 TCATACCACA AGCAAGTGAG ATTGGCTAAG GAGAAAGCTT CAAAAGTTGT
7501 CGGTGTCATG TGGGACTATG ATGAAGTAGC AGCTCACACG CCCTCTAAGT
7551 CTGCTAAGTC CCACATCACT GGCCTTCGGG GCACTGATGT TCGTTCTGGA
7601 GCAGCCCGCA AGGCTGTTCT GGACTTGCAG AAGTGTGTCG AGGCAGGTGA
7651 GATACCGAGT CATTATCGGC AAACAGTGAT AGTTCCAAAG GAGGAGGTCT
7701 TCGTGAAGAC CCCCCAGAAA CCAACAAAGA AACCCCAAG GCTCATCTCG
7751 TACCCCCACC TTGAAATGAG ATGTGTTGAG AAGATGTACT ACGGTCAGGT
7801 TGCTCCTGAC GTAGTTAAAG CTGTCATGGG AGATGCGTAC GGGTTTGTAG
7851 ATCCACGTAC CCGTGTCAAG CGTCTGTTGT CGATGTGGTC ACCCGATGCA
7901 GTCGGAGCCA CATGCGATAC AGTGTGTTTT GACAGTACCA TCACACCCGA
7951 GGATATCATG GTGGAGACAG ACATCTACTC AGCAGCTAAA CTCAGTGACC
8001 AACACCGAGC TGGCATTCAC ACCATTGCGA GGCAGTTATA CGCTGGAGGA
8051 CCGATGATCG CTTATGATGG CCGAGAGATC GGATATCGTA GGTGTAGGTC
8101 TTCCGGCGTC TATACTACCT CAAGTTCCAA CAGTTTGACC TGCTGGCTGA
8151 AGGTAAATGC TGCAGCCGAA CAGGCTGGCA TGAAGAACCC TCGCTTCCTT
8201 ATTTGCGGCG ATGATTGCAC CGTAATTTGG AAAAGCGCCG GAGCAGATGC
8251 AGACAAACAA GCAATGCGTG TCTTTGCTAG CTGGATGAAG GTGATGGGTG
8301 CACCACAAGA TTGTGTGCCT CAACCCAAAT ACAGTTTGGA AGAATTAACA
8351 TCATGCTCAT CAAATGTTAC CTCTGGAATT ACCAAAAGTG GCAAGCCTTA
8401 CTACTTTCTT ACAAGAGATC CTCGTATCCC CCTTGGCAGG TGCTCTGCCG
8451 AGGGTCTGGG ATACAACCCC AGTGCTGCGT GGATTGGGTA TCTAATACAT
8501 CACTACCCAT GTTTGTGGGT TAGCCGTGTG TTGGCTGTCC ATTTTCATGA
8551 GCAGATGCTC TTTGAGGACA AACTTCCCGA GACTGTGACC TTTGACTGGT
8601 ATGGGAAAAA TTATACGGTG CCTGTAGAAG ATCTGCCCAG CATCATTGCT
8651 GGTGTGCACG GTATTGAGGC TTTCTCGGTG GTGCGCTACA CCAACGCTGA

FIG. 2F

13/18

8701 GATCCTCAGA GTTTCCTCAAT CACTAACAGA CATGACCATG CCCCCCTGC
8751 GAGCCTGGCG AAAGAAAGCC AGGGCGGTCC TCGCCAGCGC CAAGAGGCGT
8801 GCGGGAGCAC ACGCAAAATT GGCTCGCTTC CTTCTCTGGC ATGCTACATC
8851 TAGACCTCTA CCAGATTTGG ATAAGACGAG CGTGGCTCGG TACACCACTT
8901 TCAATTATTG TGATGTTTAC TCCCCGGAGG GGGATGTGTT TGTTACACCA
8951 CAGAGAAGAT TGCAGAAGTT TCTTGTGAAG TATTTGGCTG TCATTGTTTT
9001 TGCCCTAGGG CTCATTGCTG TTGGATTAGC CATCAGCTGA ACCCCCAAAT
9051 TCAAAATTAA CTAACAGTTT TTTTTTTTTT TTTTTTTTAG GGCAGCGGCA
9101 ACAGGGGAGA CCCCAGGCTT AACGACCCCG CCGATGTGAG TTTGGCGACC
9151 ATGGTGGATC AGAACCGTTT CGGGTGAAGC CATGGTCTGA AGGGGATGAC
9201 GTCCCTTCTG GCTCATCCAC AAAAACCGTC TCGGGTGGGT GAGGAGTCCT
9251 GGCTGTGTGG GAAGCAGTCA GTATAATTCC CGTCGTGTGT GGTGACGCCT
9301 CACGACGTAC TTGTCCGCTG TGCAGAGCGT AGTACCAAGG GCTGCACCCC
9351 GGTTTTTGTT CCAAGCGGAG GGCAACCCCC GCTTGGAATT AAAAACT

FIG. 2G

14/18

1 MPVISTQTSP VPAPRTRKNK QTQASYPVSI KTSVERGQRA KRKVQRDARP
51 RNYKIAGIHD GLQTLAQAAL PAHGWGRQDP RHKSRNLGIL LDYPLGWIGD
101 VTTHTPLVGP LVAGAVVRPV CQIVRLLEDG VNWATGWFGV HLFVVCLLSL
151 ACPCSGARVT DPDTNTTILT NCCQRNQVIY CSPSTCLHEP GCVICADECW
201 VPANPYISHP SNWTGTDSFL ADHIDFVMGA LVTCDALDIG ELCGACVLVG
251 DWLVRHWLIH IDLNETGTCY LEVPTGIDPG FLGFIGWMAG KVEAVIFLTK
301 LASQVPYAIA TMFSSVHYLA VGALIYYASR GKWYQLLLAL MLYIEATSGN
351 PIRVPTGCSI AEFCSPLMIP CPCHSYLSEN VSEVICYSPK WTRPVTLEYN
401 NSISWYPYTI PGARGCMVKF KNNTWGCCR I RNVPSYCTMG TDAVWNDTRN
451 TYEACGVTPW LTTAWHNGSA LKLAILQYPG SKEMFKPHNW MSGHLYFECS
501 DTPIVYFYDP VNSTLLPPER WARLPGTPPV VRGSWLQVPQ GFYSDVKDLA
551 TGLITKDKAW KNYQVLYSAT GALSLTGVT KAVVLILLGL CGSKYLILAY
601 LCYLSLCFGR ASGYPLRPVL PSQSYLQAGW DVLSKAQVAP FALIFFICCY
651 LRCRLRYAAL LGFVPMAAGL PLTFFVAAAA AQP DYDWWVR LLVAGLVLWA
701 GRDRGHRIAL LVGPWPLVAL LTLLHLVTPA SAFDTEIIGG LTIPPVVALV
751 VMSRFGFFAH LLPRCALVNS YLWQRWENWF WNVTLRPERF FLVLVCFPGA
801 TYDALVTFCV CHVALLCLTS SAASFFGTDS RVRHRMLVR LGKCHAWYSH
851 YVLKFFLLVF GENGVFFYKH LHGDVLPNDF ASKLPLQEPF FPFEGKARVY
901 RNEGRR LACG DTVDGLPVVA RLGD LVFAGL AMP PDGWAIT APFTLQCLSE
951 RGTL SAMAVV MTGIDPRTWT GTIFRLGSLA TS YMGFVCDN VLYTAHHGSK
1001 GRR LAHPTGS IHPITVDAAN DQDIYQPPCG AGSLTRCSCG ETKGYLVTRL
1051 GSLVEVNKSD DPYWCVC GAL PMAVAKGSSG APILCSSGHV IGMFTAARNS
1101 GGSVSQIRVR PLVCAGYHPQ YTAHATLDTK PTVPNEYSVQ ILIAPTGS GK
1151 STKLPLSYMQ EKYEV LVLNP SVATTASMPK YMHATYGVNP NCYFNGKCTN
1201 TGASLTYSTY GMYLTGACSR NYDVIICDEC HATDATTVLG IGKVLTEAPS
1251 KNVRLVVLAT ATPPGVIPTP HANITEIQLT DEGTIPFHGK KIKEENLKKG
1301 RHLIFEATKK HCDELANELA RKGITAVSYY RGC DISKIPE GDCV VVATDA
1351 LCTGYTGDFD SVYDCSLMVE GTCHVDLDPT FTMGVRVCGV SAIVKGQRRG
1401 RTGRGRAGIY YYVDGSCTPS GMVPECNIVE AFDAAKAWYG LSSTE AQTIL

FIG. 3A

15/18

1451 DTYRTQPGLP AIGANLDEWA DLFSMVNPEP SFVNTAKRTA DNYVLLTAAQ
1501 LQLCHQYGYA APNDAPRWQG ARLGKKPCGV LWRLDGADAC PGPEPSEVTR
1551 YQMCFTFVNT SGTAALAVGV GVAMAYLAID TFGATCVRRRC WSITSVPTGA
1601 TVAPVVDDEE IVEECASFIP LEAMVAAIDK LKSTITTTTSP FTLETALEKL
1651 NTFLGPHAAT ILAIEYCCG LVTLPDNPFA SCVFAFIAGI TTPLPHKIKM
1701 FLSTLFGGAIA SKLTDARGAL AFMMAGAAGT ALGTWTSVGF VFDMLGGYAA
1751 ASSTACLTFC CLMGEWPTMD QLAGLVYSAF NPAAGVVGVL SACAMFALTT
1801 AGPDHWP NRL LTMLARSNTV CNEYFIATRD IRRKILGILE ASTPWSVISA
1851 CIRWLHTPTE DDCGLIAWGL EIWQYVCNFF VICFNVLKAG VQSMVNIPGC
1901 PFYSCQKGYK GPWIGSGMLQ ARCPCGAELI FSVENGFAKL YKGPRTC SNY
1951 WRGAVPVNAR LCGSARPDPT DWTSLVVNYG VRDYCKYEKL GDHIFVTAVS
2001 SPNVCFTQVP PTLRAAVAVD GVQVQCYLGE PKTPWTTT SAC CYGPDGKGKT
2051 VKLPFRVDGH TPGVRMQLNL RDALETNDCN STNNTPSDEA AVSALVFKQE
2101 LRRTNQ LLEA ISAGVDTTKL PAPSIEEVVV RKRQFRARTG SLTLPPPPRS
2151 VPGVSCPESL QRSDPLEGPS NLPPSPVLQ LAMPMP LLA GECPNFFT AIG
2201 CAMTETGGGP DDLPSYPPKK EVSEWSDESW STATTASSYV TGPPYPKIRG
2251 KDSTQSAPAK RPTKKKLKGS EFSCSMSYTW TDVISFKTAS KVLSATRAIT
2301 SGFLKQ RSLV YVTEPRDAEL RKQKV TINRQ PLFPPSYHKQ VRLAKEKASK
2351 VVGVMWDYDE VAAHTPSKSA KSHITGLRGT DVRSGAARKA VLDLQKCVEA
2401 GEIPSHYRQT VIVPKEEVFV KTPQKPTKKP PRLISYPHLE MRCVEKMYYG
2451 QVAPDVVKAV MGDAYGFVDP RTRVKRLLSM WSPDAVGATC DTVCFDSTIT
2501 PEDIMVETDI YSAAKLS DQH RAGIHTIARQ LYAGGPMIAY DGREIGYRRC
2551 RSSGVYTTSS SNSLTCWLKV NAAAEQAGMK NPRFLICGDD CTVIWKSAGA
2601 DADKQAMRVF ASWMKVMGAP QDCVPQPKYS LEELTSCSSN VTSGITKSGK
2651 PYYFLTRDPR IPLGRCSAEG LGYNPSAAWI GYLIIHYPCL WVSRLAVHF
2701 MEQMLFEDKL PETVTFDWYG KNYTVPVEDL PSIIAGVHGI' EAFSVVRYTN
2751 AEILRVSQSL TDMTMPPLRA WRKKARAVLA SAKRRGGAHA KLARFLLWHA
2801 TSRPLPDL DK TSVARYTTFN YCDVYSPEGD VFVTPQRRLO KFLVKYLAVI
2851 VFALGLIAVG LAIS

FIG. 3B

16/18



neo-RepA

...GACCGTAGCAC ATG GGG CGC GCC ATG ATT GAA CAA... (SEQ. ID. NO. 4)
 M G R A M I E Q (SEQ. ID. NO. 8)

neo-RepB

...GACCGTAGCAC ATG CCT GTT ATT TCT ACT CAA ACA GGG CGC GCC ATG ATT GAA CAA... (SEQ. ID. NO. 5)
 M P V I S T Q T G R A M I E Q (SEQ. ID. NO. 9)

neo-RepC

...GACCGTAGCAC ATG CCT GTT ATT TCT ACT CAA ACA AGT CCT GTA CCT GCG CCC GGG CGC GCC ATG ATT GAA CAA... (SEQ. ID. NO. 6)
 M P V I S T Q T S P V P A P G R A M I E Q (SEQ. ID. NO. 10)

neo-RepD

...GACCGTAGCAC ATG CCT GTT ATT TCT ACT CAA ACA AGT CCT GTA CCT GCG CCC AGA ACG CGC AAG AAC AAG CAG
 M P V I S T Q T S P V P A P R T R K N K Q
 ACG GGG CGC GCC ATG ATT GAA CAA... (SEQ. ID. NO. 7)
 T G R A M I E Q (SEQ. ID. NO. 11)

FIG. 4

17/18

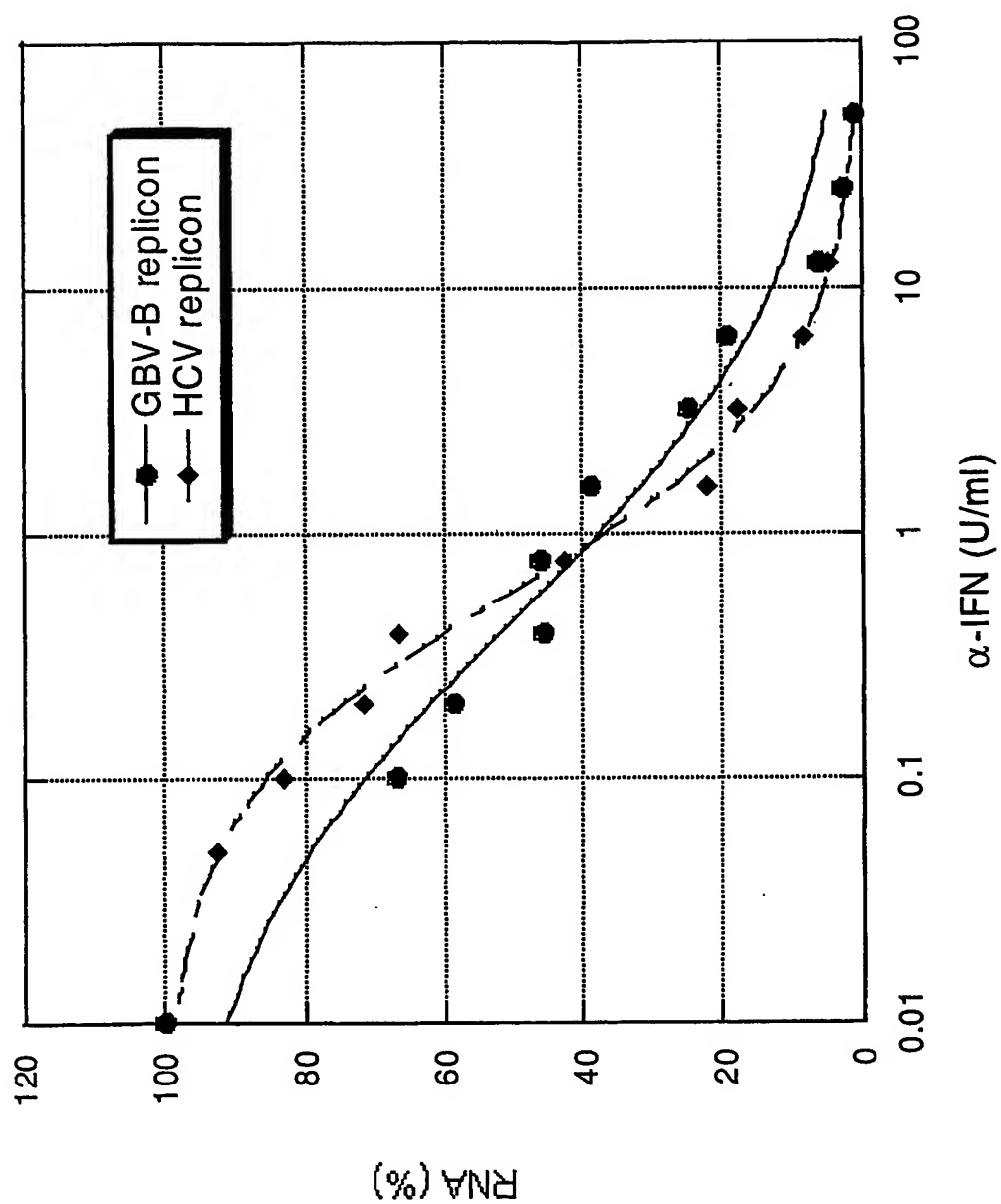


FIG. 5

NS3

HCV PTGSGKSTKVPAAYAAQGYKVLVLPNSVAATLGFGAYMSKAHGIDPNIRTGVRTITTGAP I
| | | | | : | : | : | | | | | : | :: || :: |:: | | : : : | |
GBV-B PTGSGKSTKLPLSYMQEKYEVLVLPNSVATTASMPKYMCHATYGVNPNCFNGKCNTTGAS 1290

N H G T P S R
 SC D I
 HCV LTSMLTDP SHITAETAKRRLARGSPFSLASSASQLSAPSLKATCTTRHDSPADLIEAN 2230
 GBV-B KTVKL--PFRVDGHTPGVRMQLNLRDALETND-----CNSTNNTPSDEAAVSA

HCV LLWRQEMGGNITRVESENKVVILDSFEPLQAEEDEREVSVPA--EILRRSRKFPRA---- 2282
|::||: : |:: |::| :::|: |::|:|:|
GBV-B LVFKQEL-----RRTNQLL-----EAISAGVDTTKLPAPSIEEVVVRKRQF-RARTGS

G

HCV	YSIEPLDLPQIIQR <u>L</u> HGLSAFSLHSYSPGEINRVASCLRKLGVPPLRWVRHRARSVRARL	2930
	:: : : : : : : : : : : :: :	
GBV-B	YTVPVEDLPSTIAGVHGIEAFSVVRYTNAEILRVSQSLDTMTPPLRAWRKARAVLASA	

FIG. 6